MECHANISM OF MDA5 RECOGNITION OF SHORT RNA LIGANDS AND CRYSTAL STRUCTURE OF PEPQ

A Thesis

by

TYLAN AUBREY WATTS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Pingwei Li
Committee Members,	Tatyana Igumenova
	Margy Glasner
	Bill Park
Head of Department,	Gregory Reinhart

August 2013

Major Subject: Biochemistry

ABSTRACT

The innate immune pathways that stimulate the expression of cytokines and proapoptotic factors in response to infection are triggered by the activation of the cytosolic receptors retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated gene 5 (MDA5). Activation of both receptors occurs as a result of binding to RNA. MDA5 only recognizes double stranded forms of RNA, whereas RIG-I is capable of recognizing both single and double stranded RNA. *In vivo,* MDA5 is known to be stimulated by long (>1 kb) strands of RNA, forming filaments along the phosphate backbone. However, the manner in which MDA5 can recognize the terminal end of its RNA ligand is uncertain.

I have examined the mechanism of binding of the MDA5 protein by comparing MDA5 binding to short (<18 bp) blunt RNA, 5' triphosphate RNA, and RNA with a 3' or 5' overhang. It is shown that while the MDA5 protein regulatory domain (RD) is essential for RNA recognition, the MDA5 RD only weakly recognizes short double stranded RNA ligands with overhangs or a 5' triphosphate group. The Cys951 residue was shown to disrupt stability of the MDA5 RD-RNA complex. Binding analyses were performed using a combination of SDS-PAGE, gel filtration analysis, and non-denaturing gel electrophoresis. In addition, structural data was gathered by crystallization of the MDA5 RD-RNA complex using X-ray crystallography. These results help to establish the manner in which MDA5 is regulated predominantly to the binding of long RNA ligands.

Also included in this document is structural data on the dimer form of the PepQ protein from *E. coli*. PepQ is a highly conserved proline peptidase that has a secondary activity of hydrolyzing organophosphorus triesters, toxic compounds found in many pesticides. The PepQ protein was crystallized and analyzed by X-ray diffraction. The dimer interface was clearly defined within the structure and provides insight into how the active dimer forms from the PepQ monomer.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Pingwei Li, and my committee members, Dr. Barondeau, Dr. Glasner, Dr. Park, and Dr. Igumenova, for their sage advice and guidance throughout the course of my research.

I also thank my colleagues and staff of the Biochemistry and Biophysics department of Texas A&M University, a wonderful community that made my time at Texas A&M University an enriching experience. I also want to extend my gratitude to Dr. Hays Rye and Jeremy Weaver, who provided the PepQ protein sample used in this research.

Finally, thanks go to my mother, my father, my sister, and my dear friend Dr. Ankur Gupta, who all gave me the encouragement to succeed.

The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

NOMENCLATURE

AMP-PNP	adenosine 5′-(β,γ-imido) triphosphate
CARD	caspase activation and recruitment domain
CIP	calf intestinal phosphotase
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DNA	dideoxyribonucleic acid
dsRNA	double-stranded RNA
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
EMSA	electromobility shift assay
HD	helicase domain
IPTG	isopropyl β-D-1-thiogalactopyranoside
IVT	*in vitro* transcription
LB	Luria-Bertani (growth media)
LGP2	Laboratory of Genetics and Physiology 2
MDA5	melanoma differentiation-associated gene 5
MME	monomethyl ether
MPD	(+/-)-2-methyl-2,4-pentanediol
MR	molecular replacement

PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethanesulfonylfluoride
RD	regulatory domain
RIG-I	retinoic acid-inducible gene I
RNA	ribonucleic acid
SAD	single wavelength anomalous diffraction
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion or gel filtration chromatography
ssRNA	single-stranded RNA
SUMO	small ubiquitin-like modifier
TEMED	tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
WT	wild type

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

MDA5 is an essential protein of the innate immunity pathway and serves as a cytosolic receptor for foreign dsRNA within the cell. Its activation triggers a series of signaling pathways that culminate in both apoptosis and the release of cytokines. The mechanism of how the MDA5 protein of *H. sapiens* recognizes dsRNA at its 5' end is of interest. Knowledge of the binding mechanism of MDA5 would serve to elucidate the effects certain mutations would have on MDA5's function, and help to suggest potential treatments to alleviate these effects.

The proline peptidase PepQ of *E. coli* is also of interest. While its primary function of hydrolyzing the backbone of Xaa-Pro dipeptides is significant, its secondary function of hydrolyzing organophosphorous toxins makes it a potential subject of research in biochemical waste cleanup. Here the crystal structure of the biochemically active PepQ dimer has been determined and introductory structural analysis has been performed. This paper will serve to present the cumulative experimental data collected on the mechanisms of *H. sapiens* MDA5 binding to dsRNA and on the crystal structure of *E. coli* PepQ.

CHAPTER II

RECOGNITION AND BINDING MECHANISM OF *H. sapiens* MDA5 WITH FOREIGN DOUBLE-STRANDED RNA

Introduction

During viral and bacterial infection of a cell, the first line of defense comes from the innate immune pathways. These pathways are triggered by the activation of receptors through the recognition of pathogen associated molecular patterns (PAMPs). PAMPs are unique molecules that emerge during infection, such as foreign proteins or nucleic acids. [1] Examples of receptors that bind to PAMPs include AIM2 (absent in melanoma 2), which recognizes cytosolic dsDNA, NAIP5 (NOD-like receptor family, apoptosis inhibitory protein 5), which recognizes the bacterial protein flagellin, and STING (stimulator of interferon genes), which recognizes bacterial cyclic dinucleotides. [2-4] MDA5 is a member of the RIG-I-like receptor (RLR) family of proteins, a set of PAMP receptors that respond to cytosolic RNA. [5]

RLRs include RIG-I (retinoic acid inducible gene I), MDA5, and LGP2 (Laboratory of Genetics and Physiology 2). These proteins are found in most vertebrate tissues and are localized in the cytosol. They recognize the presence of cytosolic viral RNA through their C-terminal regulatory binding domains. When activated, RIG-I and MDA5 stimulate the mitochondrial-bound IPS-1 (also known as MAVS/Cardif/VISA). The LGP2 protein in turn has been determined to act in a positive regulatory role for RIG-I and MDA5 signaling of IPS-1. [6] Once activated, IPS-1 triggers an innate

immune response through signaling pathways to phosphorylate and activate IRF (interferon regulatory factors) 3 and 7, and to activate NF-κB (nuclear factor kappa B). These pathways result in the expression of type-I interferon and proinflammatory cytokines, which assist in the cellular defense against cytolytic viruses like hepatitis C. [7] Pro-apoptotic pathways are also stimulated, helping to further arrest the progress of an active infection in the organism. [8]

Each of the RLR family of proteins contains a C-terminal regulatory domain (RD) that binds to RNA and a helicase domain (HD) that is capable of RNA-dependent ATP hydrolysis. The proteins RIG-I and MDA5 also each contain an N-terminal caspase activation and recruitment domain (CARD). [5] Upon binding to an RNA ligand, RIG-I and MDA5 CARDs become available for signaling and directly stimulate the innate immune pathway through a CARD-CARD interaction with IPS-1. [9, 10]

The RLR proteins have different structural specificities for viral RNA. RIG-I and LGP2 can recognize both single-stranded (ssRNA) and double-stranded RNA (dsRNA), while MDA5 recognizes only dsRNA. [11] Another aspect of MDA5 is that recognition of the dsRNA is not done by a single protein binding to the RNA terminus, but by a stacking series of MDA5 monomers that form a filament along the length of the RNA. [12] Structural studies of these proteins have helped to elucidate these mechanisms of RIG-I binding to the RNA terminus, and of MDA5 binding along the RNA phosphate backbone. However, preliminary studies of MDA5 with dsRNA oligos of less than 16 bp in length showed that MDA5 binds dsRNA in a 2:1 ratio. The binding was also influenced by the composition of the RNA terminus. This implies that MDA5 can bind

and recognize short RNA ligands in an alternate manner. Here I will focus on the interaction of MDA5 with the terminus of short dsRNA.

Recently, the structures of ΔCARD RIG-I and ΔCARD MDA5 bound to dsRNA (4AY2, 4GL2) have been solved. [10, 13] The structure of the RIG-I RD with dsRNA (3LRN, 3LRR) has also been solved. [14] This allows for comparison of the orientation of RNA during the initial binding to the RIG-I RD the final active complex of RIG-I and dsRNA. The structure of the MDA5 RD with dsRNA has not been solved, however. This information would help to establish how the MDA5 RD is involved with the initiation of RNA binding prior to the formation of the MDA5-RNA filament-like complex.

Objective

The manner in which MDA5 recognizes dsRNA in the cytosol is important as activated MDA5 signals an innate immune response pathway that eventually results in the expression and secretion of cytokines and the stimulation of apoptosis. Structural knowledge of MDA5 in complex with dsRNA at the point of initial interaction with RNA would be essential. This information would help to elucidate the role of MDA5 in the activation of the innate immune system in contrast with its related family members, RIG-I and LGP2. Additional studies into the stoichiometry and binding mechanism of this complex would help to further support the current research into this protein. To characterize these aspects, the following specific aims were performed.

Specific Aims

Specific aim 1: MDA5 construct expression and purification

The HD-only, RD-only, and HDRD (ΔCARD; HD and RD only) truncations of MDA5 were expressed in *E. coli*. The full length MDA5 protein was shown to be insoluble and could not be purified successfully from *E. coli.* A vector including a SUMO tag was used to improve solubility of the HD and HDRD proteins. The RD of MDA5 is essential for initial RNA binding, while the HD is involved in the subsequent conformational change to the RNA-bound active form.

Specific aim 2: Stoichiometry and stability of the MDA5 and dsRNA complex

Binding analyses of the MDA5/dsRNA complex were performed using both size exclusion chromatography (SEC) and electromobility shift assays (EMSAs). The RD, HD, and HDRD constructs of MDA5 were used. The MDA5 full length protein has a 16-18 bp footprint when bound to dsRNA [12], so dsRNA lengths were limited to less than 18 bp. This information helped establish the stoichiometry and stability of the complexes formed from these constructs.

Specific aim 3: Mechanism of the MDA5-dsRNA interaction

Mutagenesis and truncations were used to help improve the stability of the MDA5-dsRNA complex and elucidate domains essential for the binding mechanism. Binding analyses by EMSA and SEC were performed with dsRNA to determine the influence of the RNA terminus on binding by MDA5. Crystallization and X-ray

diffraction experiments were performed to attempt to obtain a crystal structure of the MDA5 RD in complex with dsRNA. A crystal structure of this complex would help to reveal the mechanisms by which MDA5 initially recognizes foreign dsRNA in the cell prior to IPS-1 signaling.

Experimental Procedures

Design of MDA5 constructs for expression

The full length (FL; 1-1025), regulatory domain (RD; residues 893-1017), helicase domain (HD; residues 301-889), helicase and regulatory domains (HDRD; residues 301-1025), HDRD with a C-terminal truncation (HDRDΔC; 301-1017), HDRD with a Hel2i-loop truncation (HDRDΔL; 301-645, 663-1025), and HDRD with both the Hel2i-loop and C-terminal truncations (HDRDΔCL; 301-645, 663-1017) were all selected for expression. The RD construct was provided by Dr. Pingwei Li in a pET22b(+) expression vector. The remaining construct designs were based upon a model of the MDA5 protein developed using the published RIG-I HD structure (PDB ID: 4A2W; Fig. 1) [9] as a template and computationally generated by SWISS-Model. [15, 16] When the MDA5 model was overlaid on the RIG-I structure, MDA5 corresponded to the RIG-I HD with few deviations (RMS = 0.081, Fig. 1). The Hel2i loop extension, not seen in the RIG-I structure, was modeled as a β-sheet. This modeling process was repeated using the RIG-I HDRD structure as a template (PDB ID: 2YKG). [17] However, while the MDA5 model again corresponded to the RIG-I structure (RMS = 0.071), the Hel2i loop was modeled in this case as a disordered loop. Secondary structure predicition algorithms Heirarchical Neural Networks [18], JPred [19], and

PsiPRED [20, 21] were all used to predict secondary structure of MDA5 and this information was correlated to the MDA5 model to design the truncations.

Fig. 1: MDA5 model overlaid on RIG-I structure (PDB ID: 4A2W). White - RIG-I HD stucture. Color - MDA5 model matching RIG-I structure. Gray - MDA5 loops not corresponding to the RIG-I structure. Arrow marks the gray Hel2i loop.

Cloning of MDA5 constructs

Forward and reverse primers were designed with *BamH*I and *Not*I restriction sites, respectively, and ordered from Integrated DNA Technology (IDT). PCR was performed using Platinum Pfx DNA polymerase (Invitrogen, USA) with 1 µL of each primer (1 mg/ml), 40 ng of cDNA template plasmid, 2 µL of MgSO4 (50 mM), 5 µL of 10x Pfx DNA polymerase amplification buffer, 5 µL of 10x Pfx DNA polymerase enhancer, 1 µL of dNTP (25 mM each of dATP, dCTP, dGTP, and dTTP) and diluted to a total volume of 50 µL by Milli-Q H_2O. The mixture was heated to 95°C for 2 min, and then cycled 35 times at 95°C for 1 min. (denaturing step), 54°C for 1 min. (annealing step), and 68°C for 2-6 min. (extension step; 2 min per kb of construct, minimum 2 min). The reaction was finished by an additional extension at 68°C for 10 min. and then cooled at 4°C for 6 min. The template plasmid DNA was eliminated using 20 U of *Dpn*I restriction enzyme (New England Biolabs, Inc., USA) at 37°C for 1 hr. The PCR product was verified by ethidium bromide-stained agarose gel and visualized by UV exposure. The product was then purified by the QIAquick PCR Purification Kit (Qiagen, Netherlands) and digested by 50 U of *BamH*I-HF and *Not*I restriction enzymes (New England Biolabs, Inc., USA) in a mixture containing 5 µL of 10x NEBuffer 4 and 5 µg bovine serum albumin, and diluted to a final volume of 50 µL by Milli-Q H_2O. 1 mg of a modified pET28a vector containing a hexahistidine tag and a SUMO tag at the N-terminus (pET28a-SUMO) was likewise digested, but using 100 U of each enzyme, 10 µL of 10x restriction enzyme buffer, and a final volume of 100 µL by Milli-Q H_2O.

The digested insert and vector fragments were purified by ethidium bromide-stained agarose gel and QIAquick Gel Extraction Kits (Qiagen, Netherlands). Ligation was carried out using 1 U of T4 DNA ligase and 1 μL of 10x T4 DNA ligase buffer (Invitrogen, USA) at 22°C for 5 hr with 400 ng of insert DNA and 200 ng of vector DNA, and then diluted to a final volume of 10 μl. All MDA5 constructs, except for the MDA5 RD, were ligated to the pET28-SUMO expression vector. Ligation success was verified by DNA sequencing. The expression vectors containing the MDA5 constructs were used to transform *Escherichia coli* BL21 (DE3) heat-shock competent cells, which were then plated on Luria Bertani (LB) agar plates containing either 100 mg/L ampicillin or 50 mg/L kanamycin. A summary of the constructs and respective vectors is provided in Table 1.

Table 1: MDA5 construct summary

Construct	Residue range	MW[†] (kD)	Expr. vector	Antibiotic
RD	301-1017	15.0	pET22b(+)	Amp
HD	301-889	67.7	pET28a-SUMO	Kan
HDRD	301-1025	83.5	pET28a-SUMO	Kan
HDRDΔL	301-645,663-1025	81.6	pET28a-SUMO	Kan
HDRDΔC	301-1017	82.6	pET28a-SUMO	Kan
HDRDΔCL	301-645,663-1017	80.7	pET28a-SUMO	Kan

[†]Proteins expressed with pET28a-SUMO contain an N-terminal SUMO tag and are 15 kD larger than listed. The MW listed is for the construct without the SUMO tag.

Cysteine mutagenesis of the MDA5 RD

Initial expression of the MDA5 RD was impeded by solubility issues with the protein. In addition, the protein construct formed a dimer during the purification process. It was hypothesized that the expressed protein formed a dimer which was mediated by a

surface disulfide bond, and that this dimer was prone to aggregation in solution. Site-directed mutagenesis on the MDA5 RD construct was performed in order to mutate the surface cysteines Cys915 and Cys951 each to serine. A mixture was prepared of 2.5 U of PfuUltra High-fidelity DNA polymerase and 5 µL of PfuUltra reaction buffer (Invitrogen, USA), 25 nmol of dNTP (dATP, dCTP, dGTP, dTTP), 125 ng each of the forward and reverse primers (Table 2), and 50 ng of the template plasmid diluted with Milli-Q H_2O to a final volume of 50 µL. The reaction mixture was heated at 95°C for 2 min., and then cycled 16 times at 95°C for 30 sec (denaturing), 54°C for 1 min. (annealing), and then 68°C for 12 min. (extension). The reaction was completed with an additional extension time of 10 min. at 68°C followed by cooling at 4°C for 10 min.

Table 2: Primers for MDA5 cysteine mutations

C915S forward primer: 5′-GCAAAAACTGCAGTGTGCTAGCCAGTTCTGGGGAAGATATCCATGTAATTG-3′
C915S reverse primer: 5′-CAATTACATGGATATCTTCCCCAGAACTGGCTAGCACACTGCAGTTTTTGC-3′
C951S forward primer: 5′-GAAAACAAAGCACTGCAAAAGAAGAGTGCCGACTATCAAATAAATGGTG-3′
C951S reverse primer: 5′-CACCATTTATTTGATAGTCGGCACTCTTCTTTTGCAGTGCTTTGTTTTC-3′

Template plasmid was eliminated by adding 20 U of *DpnI* restriction enzyme (New England Biolabs, USA) and incubating at 37°C for 1 hr. Success of the DNA

amplification was checked using an ethidium bromide stained agarose gel, visualized by UV exposure. The plasmid product was used to transform BL21 (DH5α) cells and a single-colony plasmid was isolated. Presence of the mutation within the isolated plasmid was verified by DNA sequencing. The mutated plasmid was used to transform *E. coli* BL21 (DE3) heat-shock competent cells, which were then plated on LB agar plates containing 100 mg/L ampicillin.

Small scale protein expression test of MDA5

For each construct, one colony was picked from the corresponding agar plate and used to inoculate 5 mL of LB broth. Each cell culture was incubated at 37°C for 2-4 hr with 250 rpm shaking until the OD_{600} reached approximately 0.8. To induce expression, 1 mM IPTG was introduced to 1 mL of the culture. Both the induced and non-induced cultures were incubated at 37°C for 3 hr. A 2 mL aliquot of the non-induced culture was allowed to continue incubation for an additional 2 hr, at which time it was stored at -80°C in 16% glycerol as a stock culture for use in the large scale expression. The 1 mL induced culture and 1 ml of the non-induced culture were harvested by centrifugation at 15,000 g for 1 min and their supernatants were discarded. The cell pellets were resuspended in 50 µl of lysozyme (0.2 mg/ml) and 1 ul of DNase (1 mg/ml). The cells were lysed by first flash freezing in N_2(l), then incubating at 37°C for 5 min, and then repeating this freeze-thaw cycle four more times. After lysis, the soluble and insoluble fractions were separated by centrifugation at 15,000 g for 1 min. Both fractions were then analyzed by SDS-PAGE using either an 18% gel (MDA5 RD only) or a 10% gel

(all other constructs) to determine the feasibility of expressing the respective protein in *E. coli*. For all SDS-PAGE gels, the Dual Color Protein Plus molecular weight ladder from Bio-Rad (USA) was used to determine the size of the expressed proteins.

Large scale MDA5 expression and purification

Large scale expression of the MDA5 constructs was done by first preparing a starter culture: 25-100 mL of LB broth with antibiotic (50 mg/L kanamycin or 100 mg/mL ampicillin). For each 25 mL of starter culture, a total of 25 μL of -80°C stock culture was used to inoculate. The starter culture was allowed to incubate at 37°C for 3 hr with shaking at 250 rpm. The starter culture was then used to inoculate a large scale culture, inoculating 1 L of LB broth with antibiotic (50 mg/L kanamycin or 100 mg/mL ampicillin) for each 25 mL of starter culture. Larger culture volumes were used in the case of low protein expression or solubility. For the expression of the MDA5 RD and MDA5 HDRD constructs, 100 mM $ZnSO_4$ was added to the expression cultures after inoculation. This concentration of Zn^{2+} is necessary to allow for the successful formation of the zinc-finger motif in the MDA5 RD. Once the culture reached a cell density of OD_{600}=0.8, the cultures were cooled to 15°C and then protein expression was induced by adding 400 μM IPTG. The induced culture was then incubated at 15°C for 16-20 hr.

The expression cultures were harvested by centrifugation at 4,000 g for 10 min. at 4°C, discarding the supernatant. The cell pellet was resuspended in 4°C lysis buffer (150 mL of 300 mM NaCl, 50 mM Tris, pH 8.0). The cells were lysed by sonication in an ice water bath for 10 min, using a 0.5 sec on/0.5 sec off pulse. Insoluble proteins and

cellular debris were removed through a two-step centrifugation process at 4°C: 8,000 g for 10 min. and then 16,000 g for 10 min. The insoluble pellets were discarded and the supernatant was used for subsequent purification.

The expressed MDA5 constructs were purified from the supernatant by 6xHis tag nickel affinity. The supernatant was loaded onto a Ni^{2+} Sepharose High Performance HisTrap column (GE Healthcare, USA) pre-equilibrated with 4°C lysis buffer. Non-specifically bound proteins were washed from the column using wash buffer (300 mL of 25 mM imidazole, 500 mM NaCl, 20 mM Tris, pH 7.5). The wash buffer was pre-cooled to 4°C. The high salt (500 mM) of this washing step also served to disrupt any contaminating protein or RNA complexes that may have formed during expression and subsequent cell lysis. The purified MDA5 protein was then eluted from the column using elution buffer (250 mM imidazole, 150 mM NaCl, 20 mM Tris, pH 7.5). The elution buffer was also pre-cooled to 4°C. Eluted fractions were collected in 5 mL increments and the presence of protein was detected by Bradford assay. The highest concentration fractions were pooled and final protein concentrations were measured by A_{280} and Bradford assay. To improve protein stability, 5 mM DTT was added to the purified protein samples after A_{280} measurements were completed.

To further purify the MDA5 RD construct, the eluted protein sample was concentrated to 2 ml by an Amicon Ultra-4 10 kD centrifugal filter (Millipore) at 4000 g at 4°C. The sample was then injected onto a Superdex75 16/60 HiLoad SEC column (Amersham Biosciences) pre-equilibrated with running buffer (150 mM NaCl, 20 mM Tris, pH 7.5). The sample was resolved at 1 ml/min using running buffer at a flow rate of

1.0 mL/min. and a maximum pressure of 0.35 MPa. The purified MDA5 RD protein was collected, brought to 5 mM DTT, and then placed immediately on ice before subsequent experimentation.

For the MDA5 HD construct, the SUMO tag was cleaved from the HisTrap-purified protein sample with the addition of 0.5 μM SUMO protease. Digestion was performed at 22°C for 1 hr. The digested sample was then further purified by SEC. The sample was concentrated to 2 mL by an Amicon Ultra-4 30 kD centrifugal filter (Millipore). The concentrated sample was injected onto and purified by the Superdex75 16/60 HiLoad SEC column. However, the MDA5 HDRD constructs were not further purified after the HisTrap purification until after they were prepared as protein-dsRNA complexes.

RNA preparation and purification

RNA was synthesized as blunt-end single-strand oligos by Integrated DNA Technologies (IDT, USA). Oligos were designed as GC-rich sequences of the form $GG[CG]^{N}CC$, with a total length of 10, 12, 14, or 16 bp. An additional pair of 14 bp sequences was designed with an AUAU sequence at either the 3' end (designated as 14 bp 3' overhang) or the 5' end (designated as 14 bp 5' overhang). Each RNA sample was suspended in running buffer to a concentration of 1 mM ssRNA. The diluted ssRNA was then annealed to dsRNA by placing the sample in a 95°C water bath and then immediately cooling to 22°C over 1 hr. Within this document, all dsRNA is assumed to be IDT-synthesized, blunt-ended dsRNA, unless otherwise mentioned.

RNA synthesized by *in vitro* reverse transcription was also used. The in-house RNA was 12 or 14 bp in length and contained a triphosphate at the 5' end. Removal of this cap through dephosphorylation was performed by incubation overnight at 37°C with calf intestinal phosphatase (CIP). Hereafter, *in vitro* transcribed dsRNA will be referred to as "IVT dsRNA".

MDA5 protein/RNA complex preparation and purification

Two complex forms were prepared: MDA5 RD-dsRNA and MDA5 HDRDΔCL-dsRNA. The MDA5 RD-dsRNA complexes were prepared by mixing the purified MDA5 RD protein in a 3:1 molar ratio with varied lengths of dsRNA. The RNA lengths used for each MDA5 RD-dsRNA complex included 10 bp, 12 bp, and 14 bp dsRNA. The complex was allowed to equilibrate on ice at 4°C for 30 min., after which the complex was analyzed by SEC using a Superdex200 10/30 SEC column (GE Life Sciences), pre-equilibrated with running buffer. A 100 µL aliquot was injected onto the column and the sample was resolved using running buffer at 0.4 mL/min. and a maximum pressure of 0.70 MPa. For this and all other SEC, relative protein and RNA concentrations were detected by in-line UV absorbance.

The stable complex forms were concentrated to 2 mL using an Amicon Ultra-4 10 kD centrifugal filter at 4000 g at 4℃. Purification of the MDA5 RD-dsRNA complex from unbound MDA5 RD was performed by loading the complex sample onto a HiLoad Superdex75 16/60 SEC column, pre-equilibrated with running buffer. The injected sample was resolved with running buffer at 1.0 mL/min. and a maximum pressure of

0.35 MPa. 1.5 mL fractions of the highest purity were pooled and 5 mM DTT was added to stabilize the sample. The concentration of the complex after purification was measured by Bradford assay.

The MDA5 HDRDΔCL protein (purified by HisTrap) was mixed with 12, 14, or 16 bp dsRNA at a 1.5:1 molar ratio. After dsRNA was added to the protein, the SUMO tag was removed by the addition of 0.5 μM SUMO protease. Digestion by SUMO protease and the formation of the complex occurred during incubation on ice at 4°C for 16 hr. The sample was then brought to 1% PMSF and concentrated to 2 mL by an Amicon Ultra-4 30 kD centrifugal filter. The complex was then purified from unbound dsRNA by injecting the sample onto a HiLoad Superdex200 16/60 SEC column, pre-equilibrated with running buffer. The sample was resolved with running buffer at 1.0 mL/min. and a maximum pressure of 0.35 MPa.

Binding and stoichiometry assays by SEC

The complex stoichiometry of MDA5 RD (WT or C915S/C951S) and dsRNA was analyzed using a 1-4 protein:1 dsRNA molar ratio. After the protein and dsRNA were mixed at the indicated molar ratio, the sample was allowed to equilibrate on ice at 4°C for 30 min. SEC analysis was performed by injecting a 100 μL sample onto a Superdex200 10/30 SEC column, pre-equilibrated with running buffer. The sample was resolved using running buffer at 0.4 mL/min. and a maximum pressure of 0.70 MPa.

MDA5 binding and stoichiometry by EMSA

For analyses of binding by EMSA, the concentration of dsRNA was maintained at ~5 μM with the protein maintained at 50 μM. For the stoichiometry analyses by EMSA, the molar ratio was 1-10 protein:1 dsRNA. In most cases a protein-only control and a nucleotide-only control were included for comparison. For analyses involving MDA5 HD and HDRD constructs the complex samples were equilibrated on ice for 12 hours before analyses. For analyses involving the MDA5 RD construct, the complex samples were equilibrated on ice for 30 min. before analysis. The non-denaturing polyacrylamide gel used for EMSA was prepared by mixing 1.6 mL of 40% acrylamide and bis N,N'-Methylene-bis-acrylamide (29:1 acrylamide:bis ratio, Bio-Rad, USA) with 0.8 mL of 10x Tris-boric acid buffer (890 mM Tris, 890 mM boric acid) to a final volume of 7.1 mL. To polymerize the gel, 80 μL of 10% ammonium persulfate and 10 μL of TEMED were added. Samples were loaded and resolved at 150 V for 45 min. The gel was stained by ethidium bromide and exposed to UV to visualize the dsRNA.

Crystallization, data collection and processing

The purified complexes of MDA5 RD C915S/C951S and 10, 12, or 14 bp dsRNA were concentrated to 25 mg/mL, measured by Bradford assay. The purified complexes of MDA5 HDRDΔCL and 12, 14, or 16 bp dsRNA were concentrated to 10 mg/mL, measured by Bradford assay. For the MDA5 HDRDΔCL 2 mM of a non-hydrolyzable ATP analogue, AMP-PNP, was added. Crystallization screening for all complexes was performed by the hanging-drop method, using 2 μL of complex sample

and 2 μL of screening condition over a reservoir of 0.5 mL mother liquor. The screens were set up and incubated for at least 5 days in a temperature-controlled environment at either 4°C or 16°C. The screening conditions were obtained as kits from Hampton Research, Inc. (USA): Index, Crystal Screen, Crystal Screen 2, and PEGRx 1.

The MDA5 RD C915S/C951S and 12 bp dsRNA complex was crystallized in 17.5% PEG 3350, 0.1 M succinic acid, pH 7.0 and grown for 15 days at 4°C. The complex crystal was transferred stepwise into a cryobuffer containing 30% PEG 400, 17.5% PEG 3350, 0.1 M succinic acid, pH 7.0. The crystal was then mounted in a cryo loop and flash-frozen in N_2(l). Diffraction data were collected at 100 K at the Lawrence Berkeley National Labs Advanced Light Source using a 3 x 3 CCD array at a wavelength of 1.28254 Å. The wavelength was chosen to permit single wavelength anomalous diffraction (SAD) using the Zn heavy atom of the MDA5 RD zinc-finger motif. The crystal was rotated through 120° with 1° oscillations and 5 seconds exposure per frame. Data were processed using the HKL2000 package [22]. Attempts at resolving the phasing problem by SAD and molecular replacement (MR) were attempted using the Phenix package. [23]

Results

MDA5 mutation, expression, and purification

Initial expression of the MDA5 RD construct (15 kD) was performed in a 2 L culture and purified as described. The HisTrap purification process was analyzed by an 18% SDS-PAGE before purifying by SEC. (Fig. 2A) The UV chromatogram for the wild

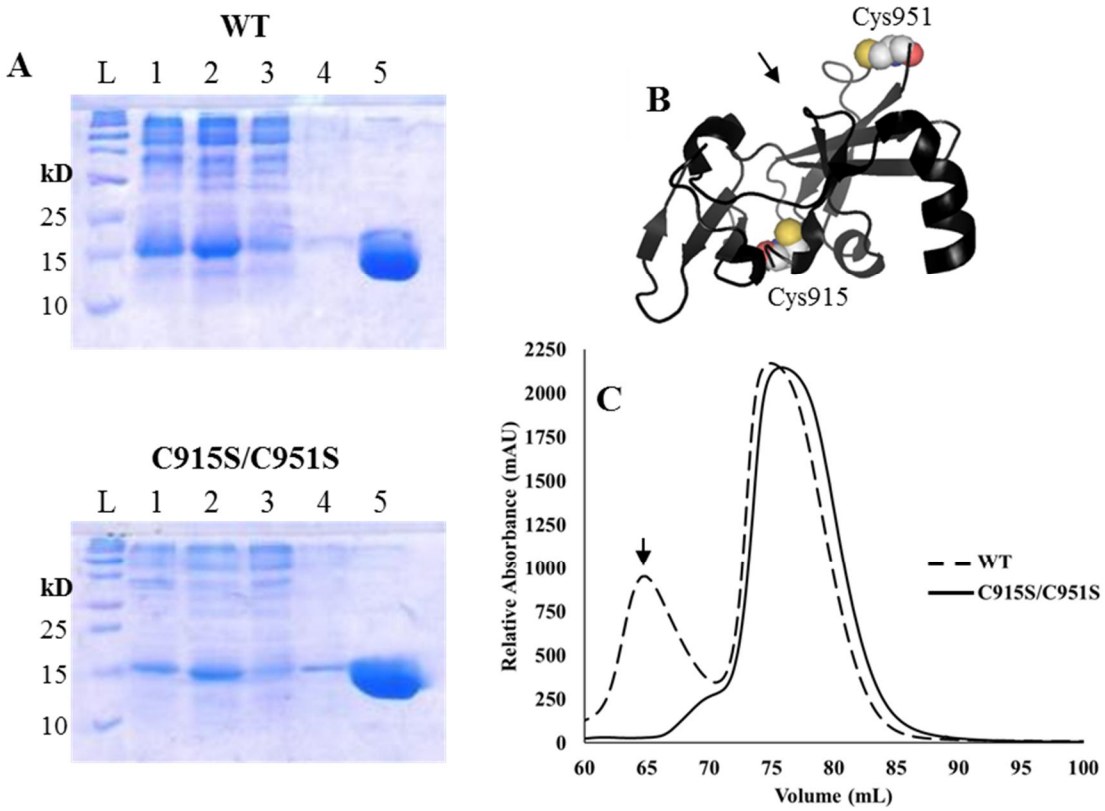

Fig. 2: MDA5 RD expression. A: SDS-PAGE analyses of HisTrap purification of the wild type and C915S/C951S mutant proteins; L – protein ladder; lanes 1,2 – insoluble fraction; lane 3 – unpurified soluble fraction; lane 4 – wash fraction; lane 5 – elute. B: MDA5 RD structure (PDB 3GA3) showing location of the mutated cysteine residues as spheres; arrow indicates RNA binding surface. C: UV chromatograms of SEC purification; arrow indicates position of the MDA5 RD WT dimer peak.

type (WT) MDA5 RD showed the presence of an unknown secondary peak that resolved to a position corresponding to the molecular weight of an MDA5 RD dimer. (Fig. 2C) This dimer formation was possibly the result of two surface cysteines forming a disulfide bond, and so site-directed mutagenesis was performed. Two surface cysteine residues candidates were identified using the MDA5 RD crystal structure and mutated to serine residues (PDB 3GA3; Fig. 2B). [24] The new MDA5 RD C915S/C951S protein

was expressed in the same manner as the WT (Fig. 2A). The contaminating dimer peak was effectively eliminated as shown by the UV chromatogram of the gel purification. (Fig. 2C) This mutant had an expression and yield of protein comparative to the WT.

The MDA5 HD construct was expressed as a SUMO-tagged protein in a 2 L culture and purified as described. The HisTrap purification process was analyzed by a 10% SDS-PAGE gel owing to the larger size of the protein (68 kD without the SUMO tag). A significant portion of the expressed protein was found in the insoluble pellet,

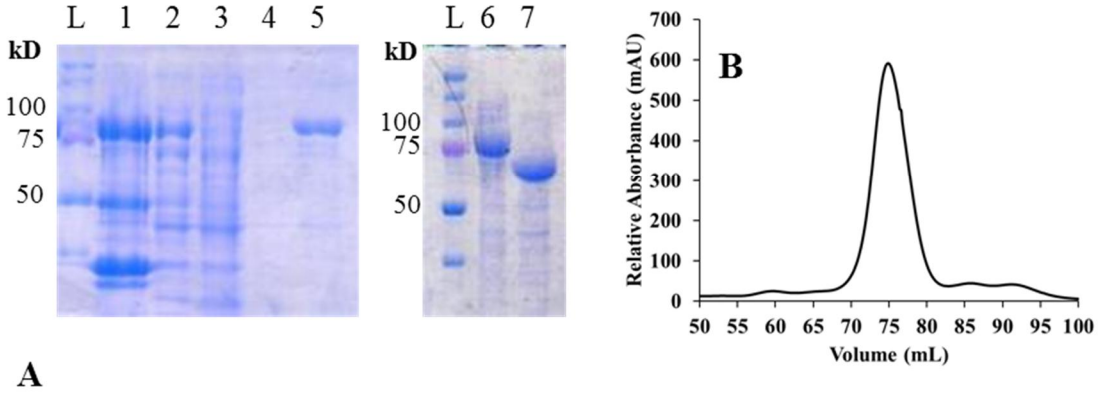

Fig. 3: MDA5 HD expression. A: SDS-PAGE of HisTrap purification and SUMO protease digestion; L – protein ladder; lane 1,2 – insoluble fraction; lane 3 – unpurified soluble fraction; lane 4 – wash fraction; lane 5 – elute; lane 6 – SUMO-tagged MDA5 HD; lane 7 – MDA5 HD after digestion by SUMO protease. B: UV chromatogram of MDA5 HD purification by SEC.

despite the presence of the soluble SUMO tag. (Fig. 3A) SUMO protease digestion of the soluble HisTrap-purified MDA5 HD was performed at 22°C for 1 hr. Purification by HisTrap, digestion by SUMO protease, and purification by SEC (Fig. 3B) were all completed the same day that the expression culture was harvested. The purified protein was stable and remained soluble for several days on ice in the presence of 5 mM DTT.

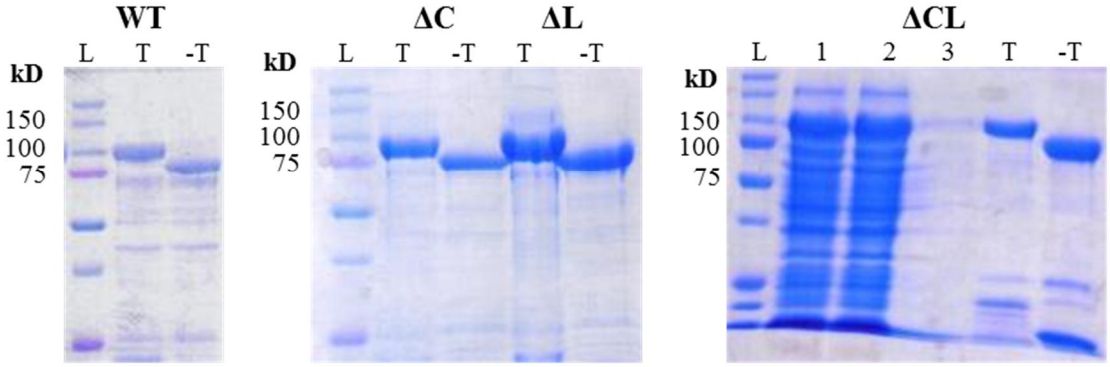

Fig. 4: MDA5 HDRD expression. SDS-PAGE of SUMO digestion of MDA5 HDRD constructs and HisTrap purification of MDA5 HDRDΔCL. For all constructs: L – protein ladder; T – SUMO-tagged protein; -T – protein after removal of SUMO tag. For ΔCL construct: 1 – insoluble fraction; 2 – soluble fraction; 3 – wash fraction.

The MDA5 HDRD constructs (each ~81-84 kD without SUMO tag) were expressed as SUMO-tagged proteins. The WT, ΔC, ΔL and ΔCL constructs were each purified by HisTrap, digested by SUMO protease, and then analyzed by 10% SDS-PAGE gel. (Fig. 4) The construct with the highest purity and stability was determined to be the ΔCL truncation. Initial purification of the protein by SEC was problematic as the MDA5 HDRDΔCL protein would aggregate and degrade within a few hours after His-Trap purification unless the sample was stabilized by binding to RNA prior to digestion by SUMO protease. In addition, the protein also suffered from proteolytic degradation prior to purification by SEC. To mitigate this, 1% PMSF, a serine protease inhibitor, was added immediately after SUMO protease digestion.

Stoichiometry of MDA5 to dsRNA

As the RIG-I and LGP2 RD protein complexes with dsRNA have a stoichiometry of 2 protein:1 dsRNA, it was hypothesized that the same would be true of the MDA5 RD. [11, 14] Binding analyses of the MDA5 RD WT protein with 14 bp dsRNA were

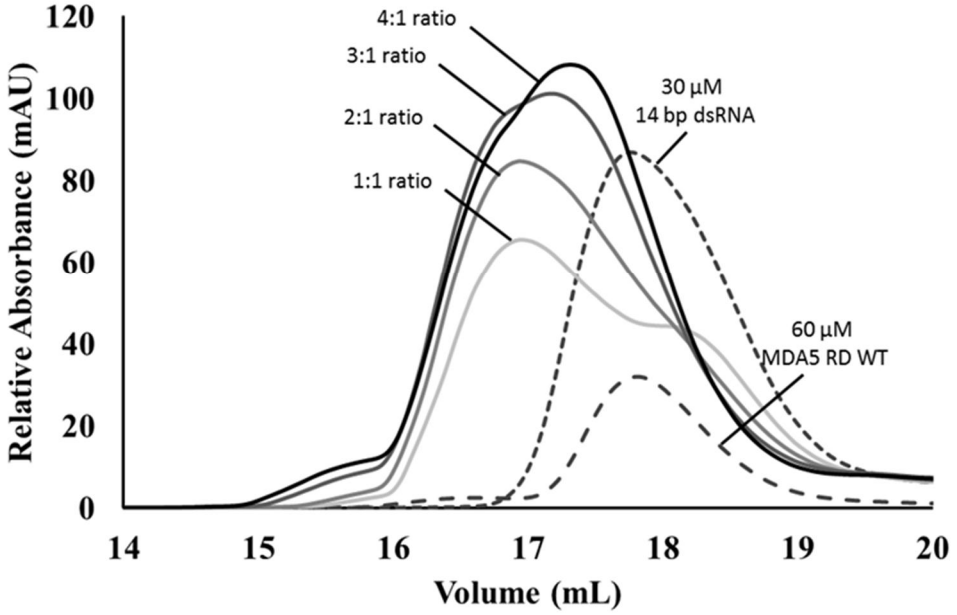

Fig. 5: MDA5 RD WT binding to 14 bp blunt dsRNA. With the exception of the protein-only control, 30 µM dsRNA was constant for each sample. The protein-only control contained 60 µM MDA5. The protein:dsRNA molar ratio is noted for each analyzed complex sample.

performed by SEC using a molar ratio of 1-4 protein:1 dsRNA. (Fig. 5) The UV chromatograms show that the binding stoichiometry is approximately 2 protein:1 dsRNA. When the analyses were repeated with the MDA5 RD C915S/C951S mutant, the binding stoichiometry was similar, but the binding stability was greatly increased. (Fig. 6) This binding change may be due to the MDA5 RD no longer forming a stable dimer, or it may be due to the cysteine mutant having an influence on the binding site.

Cys951 is in close proximity to the RNA binding surface (Fig. 2B) and may be responsible for this effect.

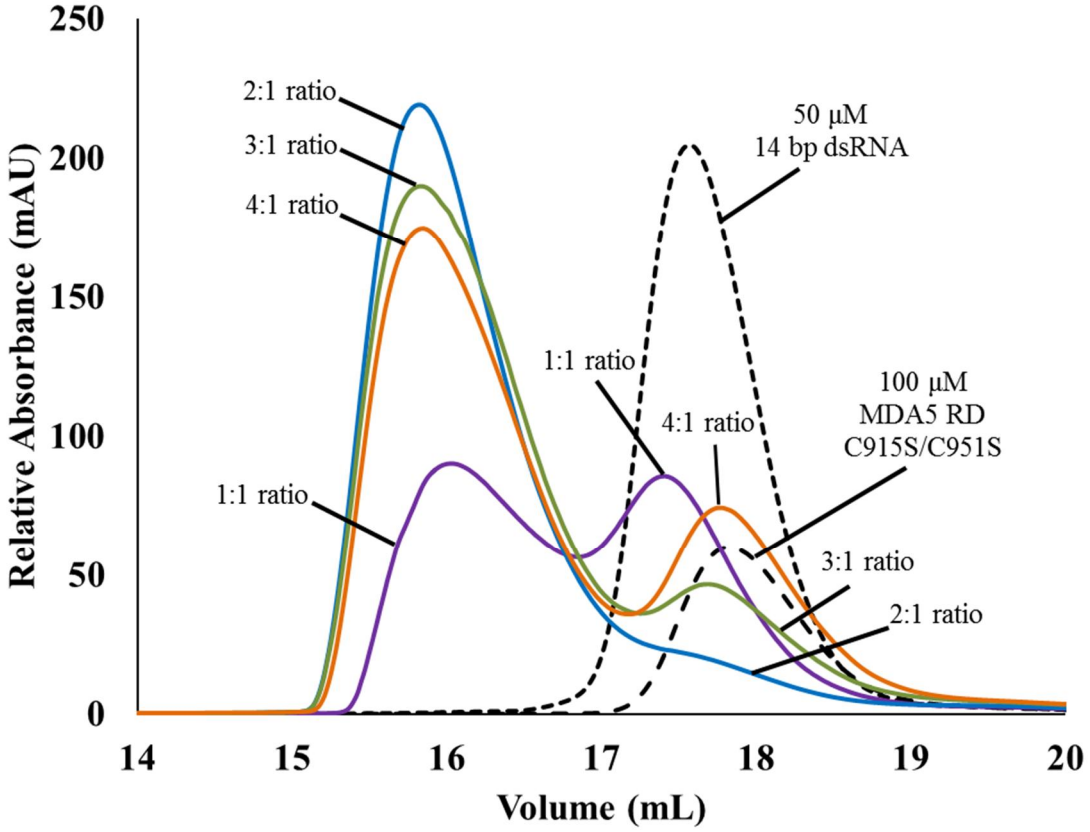

Fig. 6: MDA5 RD C915S/C951S binding to 14 bp blunt dsRNA. UV chromatograms generated from SEC of MDA5 RD C915S/C951S with 14 bp dsRNA. With the exception of the protein-only control, 50 µM dsRNA was constant for each sample. The protein-only control contained 100 µM MDA5. The protein:dsRNA molar ratios are noted for each analyzed complex sample.

For the MDA5 HDRD constructs, EMSA was used instead of SEC analysis to analyzed stoichiometry. (Fig. 7) Due to protein instability, the SUMO tag was not removed from MDA5 HDRD prior to analysis. Based upon the data, the point at which all of the RNA is bound to the MDA5 protein is between 5-10 protein:1 dsRNA. This

high ratio is supported by Piesley et al. who showed that at low concentrations MDA5 has a low affinity for short dsRNA. [25] In addition, the EMSA shows that the protein-

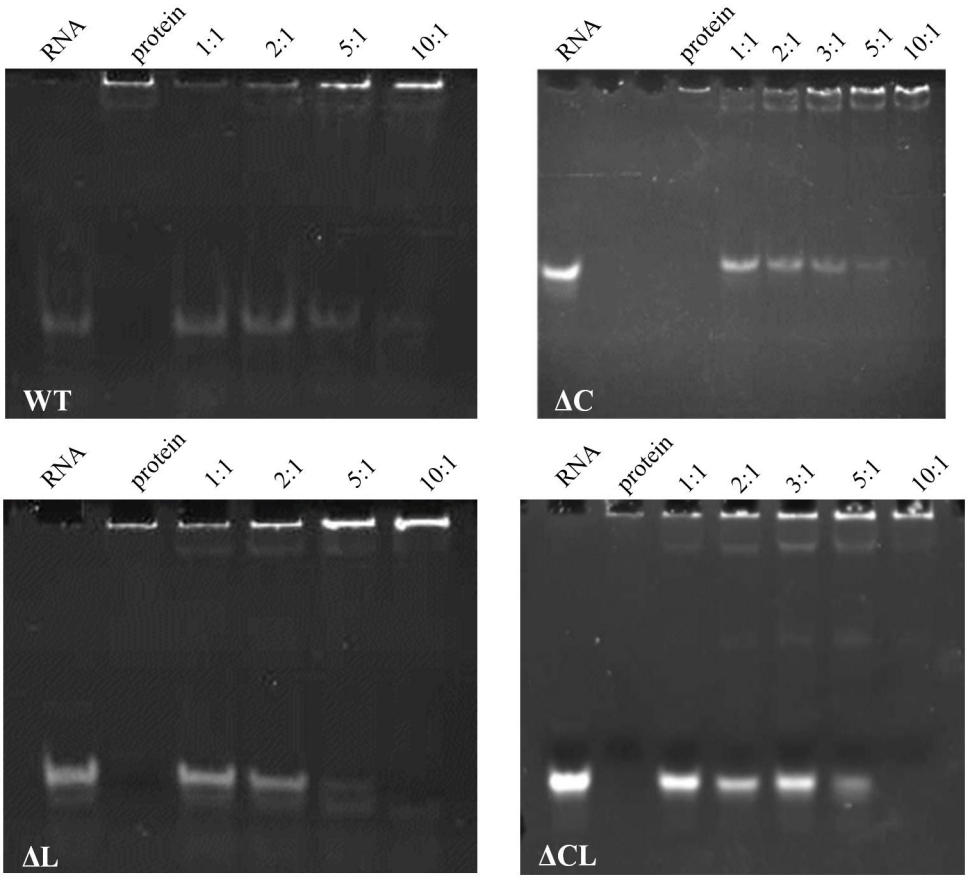

Fig. 7: EMSA of MDA5 HDRD constructs complexed with 12 bp blunt dsRNA. RNA and protein lanes are controls. All listed molar ratios are in the format protein:dsRNA.

only control samples contain some nucleic acid contaminant. This is possibly RNA from *E. coli* that was bound to the MDA5 protein during expression or lysis and not removed by the 0.5 M NaCl wash of the HisTrap purification. As a result, this data was only used to determine the existence of RNA binding, not the stoichiometry of the complex.

MDA5 binding to dsRNA

Analysis of the mechanism of MDA5 RD binding to dsRNA was examined. The binding of MDA5 RD (WT and C915S/C951S mutant) to dsRNA with a 5' overhang, a 3' overhang, a 5' triphosphate group, or a blunt end were tested by EMSA. (Fig. 8) This analysis was also performed by SEC for the C915S/C951S mutant. (Fig. 9) As expected,

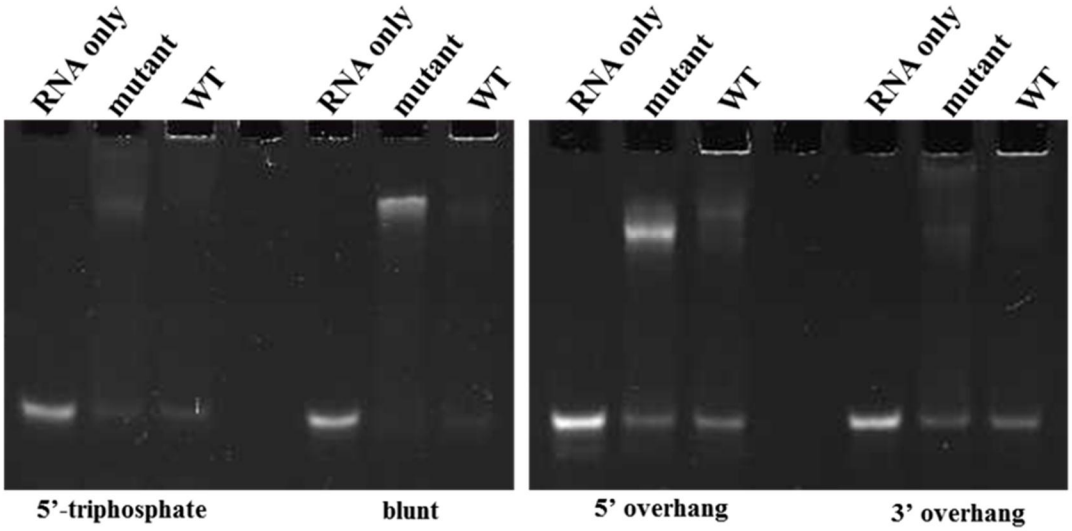

Fig. 8: Effect of C915S/C951S mutation on MDA5 RD ligand recognition. EMSA of four forms of 14 bp dsRNA complexed with MDA5 RD WT and MDA5 RD C915S/C951S constructs. WT and mutant headers designate the WT complex and C915S/C951S complex respectively. Complexes are formed from a molar ratio of 10 protein:1 dsRNA.

the tests showed that the strongest binding occurred with blunt dsRNA. Binding was seen with the 5' triphosphate dsRNA, which was confirmed by SEC. Binding was also seen with the 3' overhang dsRNA, but no binding was seen by SEC analysis. The 5' overhang dsRNA also showed binding with MDA5 RD by EMSA, but as with the 3' overhang, no binding of this oligo was seen using SEC. The C915S/C951S mutant shows a different behavior on the EMSA compared to the WT. Complex formation with

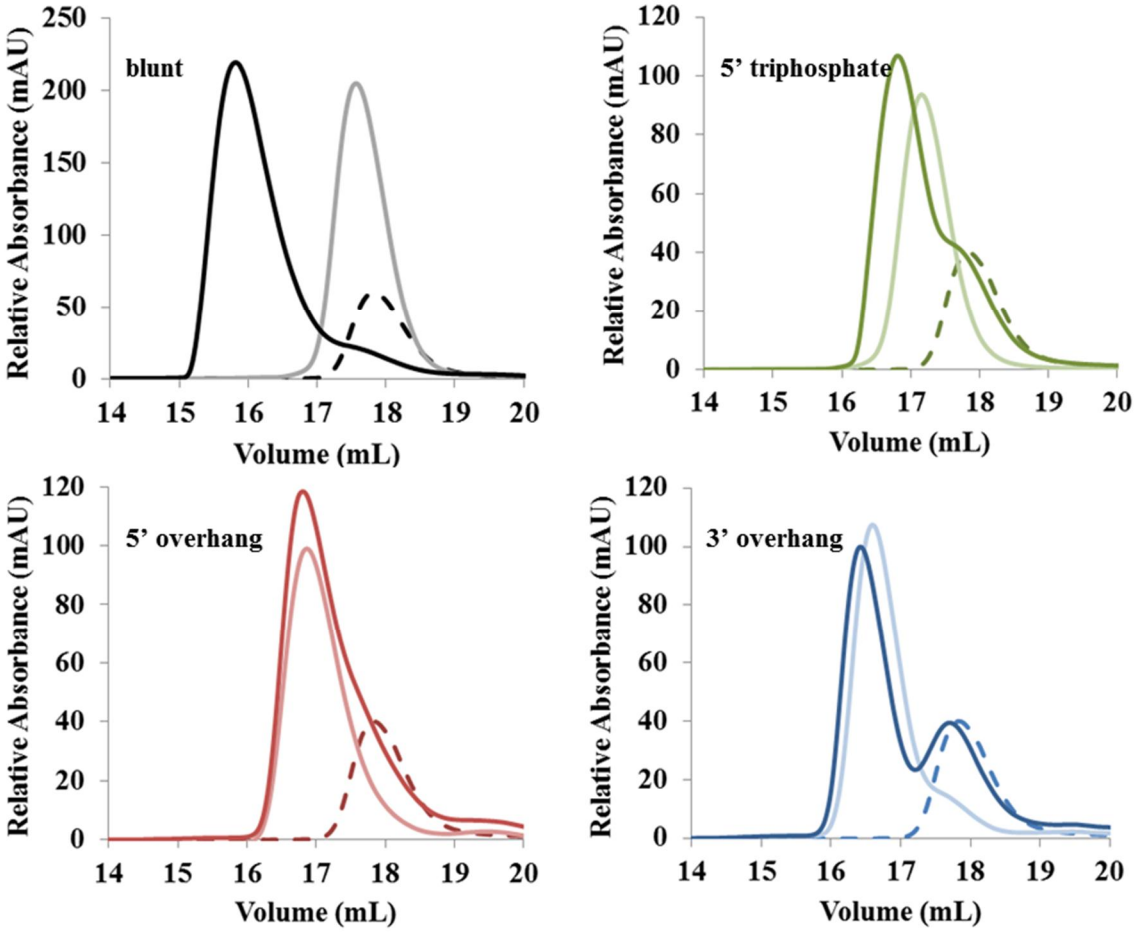

Fig. 9: Effect of dsRNA ends on the binding stability of MDA5 RD C915S/C951S. UV chromatogram results of SEC on MDA5 RD C915S/C951S protein mixed with alternate forms of dsRNA. Light line – dsRNA control; dashed line – protein control; solid dark line – 2:1 complex.

the blunt-end, 5' overhang, and 5' triphosphate forms of dsRNA is more distinct for the C915S/C951S mutant compared to the WT. This difference was also seen by SEC during the prior stoichiometry experiments.

The SEC method relies upon separation by high pressure flow from a low volume into a high volume matrix composed of porous beads using a buffer composed of moderate salinity and pH (150 mM NaCl, 20 mM Tris, pH 7.5). The EMSA method

instead applies a voltage gradient to draw the sample through a porous gel and contains lower salinity (89 mM Tris, 89 mM boric acid). The relative concentration of the complex sample within the EMSA gel (0.45 μM) is also higher than within the SEC matrix (0.16 μM). The higher salinity of the SEC buffer coupled with the lower concentration of complex may impair the binding of MDA5 RD with the 3' and 5' overhang dsRNA.

The binding of 5' triphosphate dsRNA to MDA5 was further investigated using the HDRDΔCL construct and the RD C915S/C951S mutant. For this experiment, the SUMO tag was not removed from the HDRDΔCL. A sample of 12 bp IVT dsRNA and 14 bp IVT dsRNA were each dephosphorylated by CIP. An EMSA was then used to test and compare the binding capacity of MDA5 HDRDΔCL to 12 bp dsRNA with and without a 5' triphosphate cap. (Fig. 10A) Concentrations of RNA were constant for each lane. The 5' triphosphate group of the IVT dsRNA appears to inhibit binding by the HDRDΔCL protein, but after CIP-digestion the binding behavior resembled that of blunt dsRNA. Unfortunately, all lanes contained several contaminating bands, possibly the result of contaminating bacterial RNA, making it difficult to make a positive conclusion.

The impact of the 5' triphosphate group on dsRNA binding with the MDA5 RD C915S/C951S construct was analyzed using SEC. (Fig. 10B and 10C) The 14 bp blunt dsRNA complex resolved at 16.1 mL while the RNA-only control resolved at 17.8 mL. The complex using CIP-digested 14 bp IVT dsRNA resolved at 16.5 mL, while the RNA-only control resolved at 17.3 mL. With uncut 14 bp IVT dsRNA, the complex

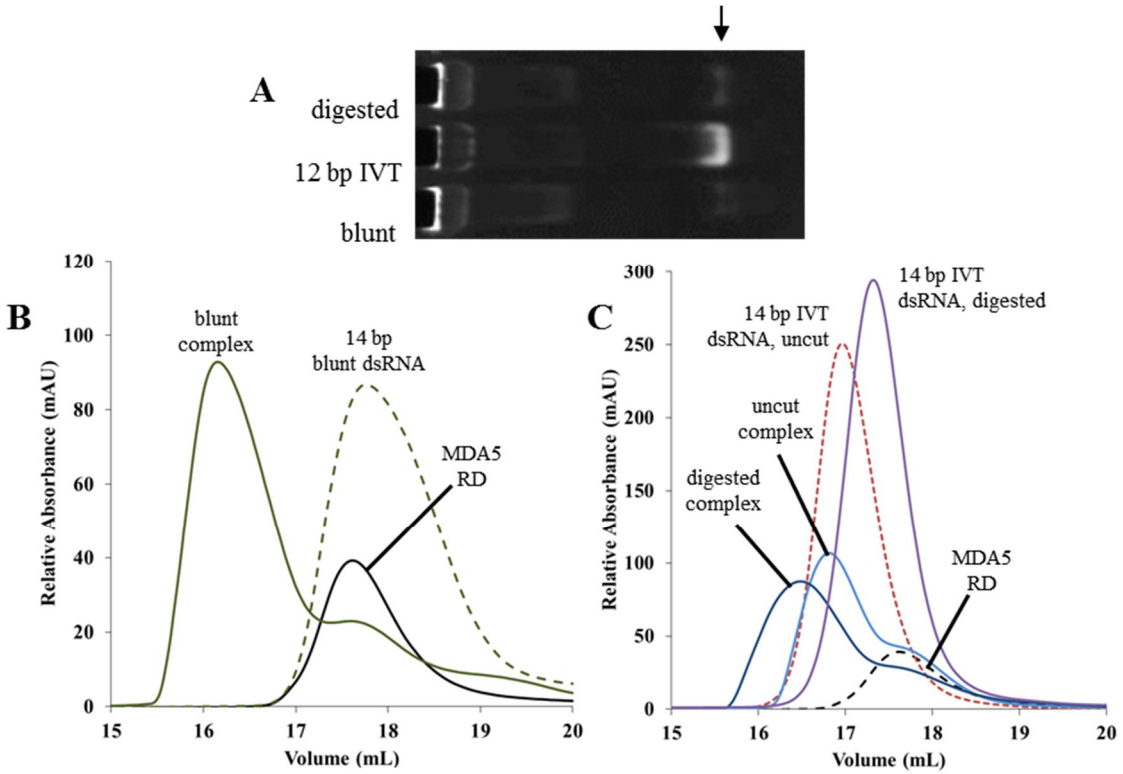

Fig. 10: MDA5 binding to IVT dsRNA before and after CIP digestion. A: UV chromatogram of SEC analysis of MDA5 RD C915S/C951S in complex with 14 bp blunt dsRNA. B: UV chromatogram of SEC analysis of MDA5 RD C915S/C951S in complex with 14 bp IVT dsRNA before and after digestion by CIP. C: EMSA of MDA5 HDRDΔCL in complex with 12 bp dsRNA; lane 1 – blunt dsRNA, lane 2 – IVT dsRNA with an undigested 5' triphosphate group, lane 3 – IVT dsRNA after removal of the 5' triphosphate group by CIP digestion. Arrow indicates unbound dsRNA.

resolved at 16.8 mL but the RNA-only control resolved at 17.0 mL. The significant difference in resolution positions for the uncut IVT dsRNA and the CIP-digested sample suggest that the 5' triphosphate group was successfully removed.It is clear from the chromatogram comparison that the presence of the 5' triphosphate impairs binding of the MDA5 RD C915S/C951S complex with dsRNA. Unfortunately, complexes prepared using this relatively less expensive RNA source would not crystallize, suggesting that crystallization of this complex is more sensitive than first suspected.

Binding analysis was also performed with the MDA5 HD construct to see if the HD was directly involved with RNA binding. As before, the RNA used was 12 bp dsRNA, an oligo known to bind strongly to both the RD and HDRD constructs. However, no binding of dsRNA with the MDA5 HD construct was seen by either EMSA or SEC analysis. (Fig. 11)

It is known that stimulation and binding of the full length MDA5 protein to dsRNA *in vivo* is independent of the presence of a 5' triphosphate end. Binding behavior is instead dependent upon the length of the foreign RNA. [26] Based upon the data shown here, it appears that the MDA5 protein directly interacts with the 5' end of dsRNA even if the length of the RNA is below the 0.5-7 kb stability threshold. [17] The data shows that phosphate groups impair MDA5 binding when there is interaction with

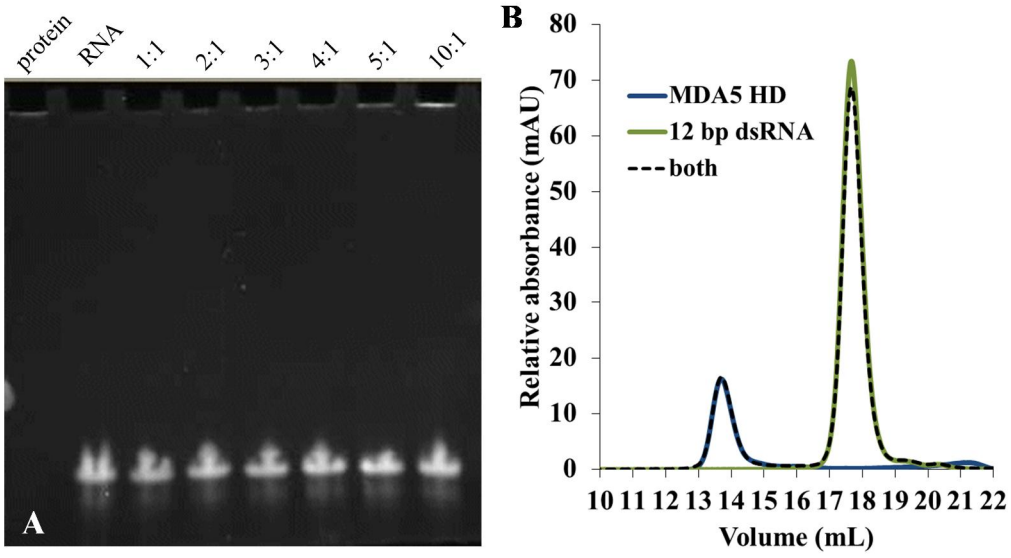

Fig. 11: MDA5 HD binding to 12 bp dsRNA. A: EMSA of MDA5 HD in complex with 12 bp dsRNA; molar ratios are in the format protein:dsRNA. B: UV chromatogram of SEC analysis of MDA5 HD C915S/C951S in complex with 12 bp dsRNA. No interaction of MDA5 HD with dsRNA was observed in either analysis.

the 5' end of dsRNA. In the case of dsRNA with 5' overhangs or 3' overhangs, ligand concentration and ionic strength of the solution may have an effect on whether the complex can readily form.

MDA5 HDRD crystal screening

Crystallization screening of the complex of MDA5 HDRD with dsRNA was performed using 12 bp dsRNA and the ΔCL construct. The complex was prepared by mixing the expressed SUMO-tagged protein with 12 bp dsRNA in a 1.5 protein:1 dsRNA molar ratio. The complex mixture was incubated overnight at 4°C on ice after adding 5 mM DTT and 0.5 μM SUMO protease. The sample was tested by 10% SDS-PAGE to confirm SUMO protease digestion. Immediately before the sample was concentrated for SEC purification, 1% PMSF was added to the sample. The complex was purified by SEC and the purified complex was collected as 1.5 mL fractions. These fractions were examined by 10% SDS-PAGE and denaturing polyacrylamide gel. Both the MDA5 HDRDΔCL and the 12 bp dsRNA were verified to be present in the complex peak. (Fig. 12) The purified complex was concentrated to 10 mg/mL for crystallization screening at 4°C.

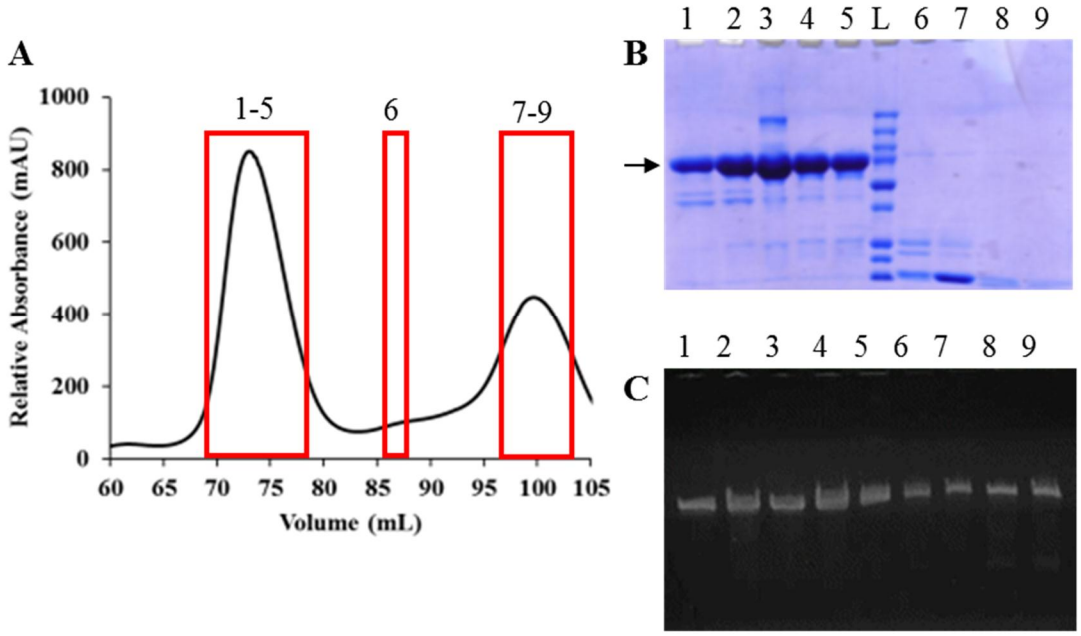

Fig. 12: Purification of MDA5 HDRDΔCL complex with 12 bp dsRNA. A: UV chromatogram of purification by SEC; ranges listed above red boxes correspond to fractions analyzed in B and C; fractions 1-5 are the complex, 6-9 is unbound dsRNA. B: 10% SDS-PAGE analysis of fractions; L – protein ladder; arrow indicates the MDA5 HDRDΔCL protein. C: denaturing polyacrylamide gel analysis of fractions, visualized by UV.

To optimize the crystallization, AMP-PMP, a non-hydrolyzable ATP analogue, was added at 2 mM. It was hypothesized that the same inhibitory effect of this molecule that allowed crystallization of the RIG-I HDRD complex with dsRNA would provide the same benefit for crystallization of the MDA5 HDRDΔCL complex with dsRNA. No crystals formed during the initial screen at 4°C. Additional screens at 16°C produced similar results. MDA5 HDRDΔCL-dsRNA complexes using 14 bp dsRNA and 16 bp dsRNA were also screened, but no crystallization occurred.

MDA5 RD crystallization optimization, diffraction, and data analysis

A complex sample was prepared using the MDA5 RD C915S/C951S construct and 14 bp dsRNA. The protein and dsRNA were mixed in a 3 protein:1 dsRNA molar ratio and allowed to incubate on ice for 30 min. The complex was then purified by SEC to remove any unbound protein. (Fig. 13A) The purified complex was concentrated to 25

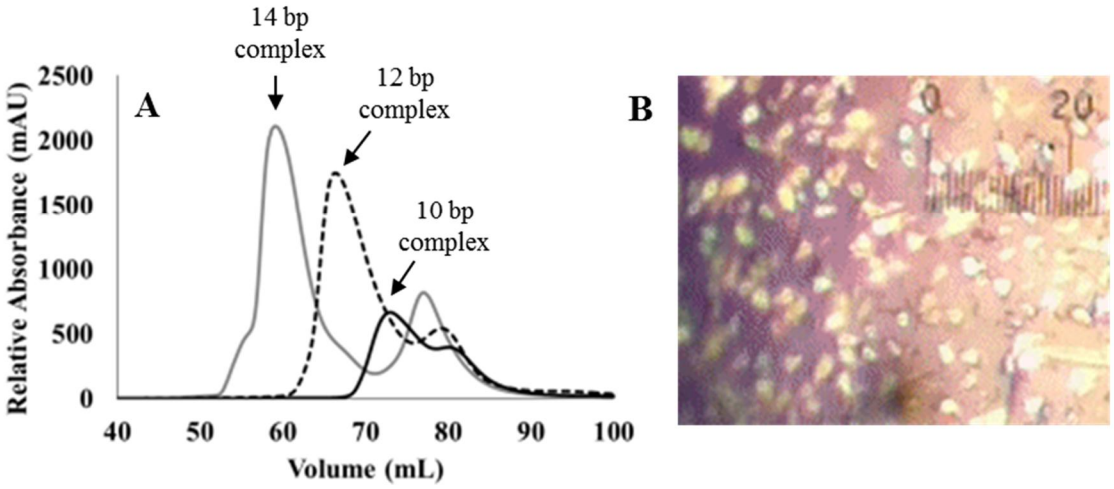

Fig. 13: Purification and crystallization of MDA5 RD C915S/C951S with dsRNA. A: UV chromatogram of purification by SEC of MDA5 RD C915S/C951S complexed with three lengths of blunt dsRNA – 10 bp, 12 bp, and 14 bp; arrows indicate the positions of the complex peaks. B: Optimized crystals of the MDA5 RD C915S/C951S and 14 bp dsRNA complex. Scale is 63 μM per unit.

mg/mL and used for crystallization screening. The screens were performed by the hanging-drop method at 4°C. Needle-like crystals formed in 22% (w/v) polyacrylic acid Na salt, 5100, 20 mM $MgCl_2$, 0.1 M HEPES, pH 7.5. The crystals were not discrete, and so optimization was performed using a lower range of precipitant. This process produced single crystals, but the optimized crystals (Fig. 13B) did not grow large enough to be usable for diffraction. To find an alternate crystallization condition potentially more

receptive to optimization, shorter dsRNA oligos were used for complex formation: 10 bp and 12 bp dsRNA. These complex samples were prepared and purified in the same manner as the 14 bp dsRNA complex. (Fig. 13A)

Screens using the purified MDA5 RD C915S/C951S and 10 bp dsRNA complex produced three crystals (Fig. 14), each large enough to be used in X-ray diffraction analysis. A 1.3 mm x 2.8 mm x 0.7 mm crystal formed in 45% MPD, 0.2 M Ammonium acetate, 0.1 M BIS-Tris, pH 5.5. A separate 0.4 mm x 2.0 mm x 0.6 mm crystal formed in 30% PEG 400, 0.2 M Sodium citrate tribasic dihydrate, 0.1 M Tris, pH 8.5. A third 2.6 mm x 2.4 mm x 0.6 mm crystal formed in 0.2 M NaCl, 10% PEG 6000. All of the

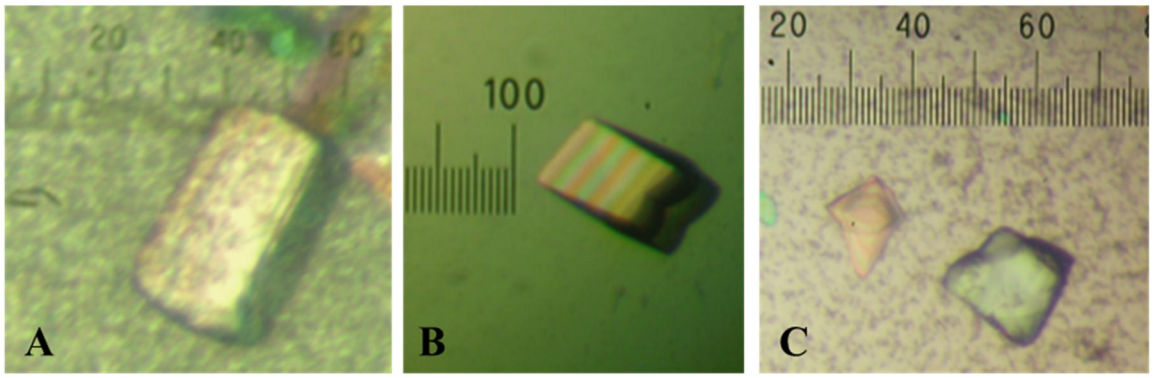

Fig. 14: RNA crystals formed from a 10 bp dsRNA complex sample. A: crystal formed in 45% v/v MPD, 0.2 M Ammonium acetate, 0.1 M BIS-Tris, pH 5.5. B: crystal formed in 30% v/v PEG 400, 0.2 M Sodium citrate tribasic dihydrate, 0.1 M Tris, pH 8.5. C: crystal formed in 0.2 M NaCl, 10% PEG 6000. Scale shown is 63 μM per unit.

crystals formed after 5 days incubation at 4°C. The crystals were washed in their respective original conditions and then flash frozen in N_2(l). The crystals were analyzed at 180 K by X-ray diffraction using a Raxis IV++ detector mounted on a Rigaku Micromax-007HF generator. The collected data were processed and it was determined

that in all three cases the crystals were composed exclusively of dsRNA and did not contain any of the MDA5 RD C915S/C951S protein.

The purified MDA5 RD C915S/C951S and 12 bp dsRNA complex sample was screened in the same manner as the 10 bp and 14 bp dsRNA complexes. Crystals formed in 15% PEG 3350, 0.1 M succinic acid, pH 7.0 at 4°C after 6 days. (Fig. 15A) To verify the presence of protein and RNA in the crystal, a sample crystal was washed in its

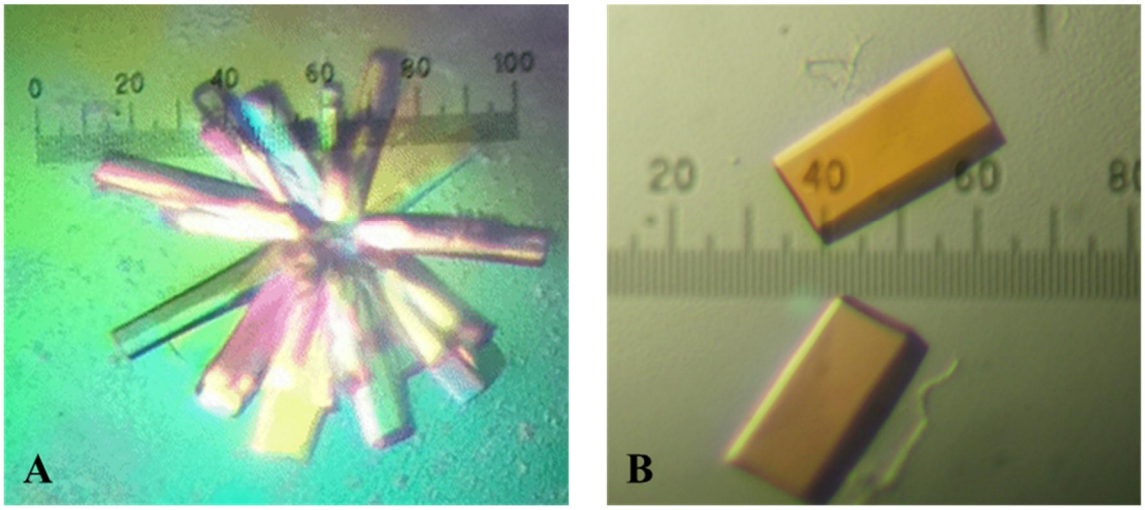

Fig. 15: Crystals formed from MDA5 RD C915S/C951S in complex with 12 bp dsRNA. A: crystal formed in the initial optimization screen. B: optimized crystal. Scale shown is 63 μM per unit.

original condition and then analyzed by both 18% SDS-PAGE and denaturing polyacrylamide gel. Visualization of the gels by Coomassie blue and ethidium bromide staining revealed the presence of both protein and RNA in the washed crystal. Optimization of this crystallization condition produced a crystal suitable for data collection. The crystal was grown in 17.5% PEG 3350, 0.1 M succinic acid, pH 7.0 at

4°C for 15 days to a final size of 0.7 mm x 1.7 mm x 0.4 mm. (Fig. 15B) The crystal was transferred stepwise to a cryobuffer containing the original condition plus 30% PEG 400. The crystal was then flash frozen in N_2(l) and analyzed at the Lawrence Berkeley National Labs synchrotron. The wavelength chosen for diffraction was 1.28255 Å. This wavelength was determined by fluorescence scanning to have the highest signal-to-noise ratio for Zn-SAD. Diffraction data were collected and processed using the HKL2000 package.

The Phenix package was used to solve the phasing problem. [23] The crystal was determined to have severe pseudo merohedral twinning, which was the case for all of the crystals that were grown using this condition and analyzed by X-ray diffraction. This indicates that the twinning is likely an artifact of the crystal's growth pattern and not the freezing process. The data was determined to have a 0.45 twin fraction using the –h, -k, +l law. In addition, the low overall signal strength of the Zn anomalous signal prevented resolution of the structure by SAD, and despite numerous attempts the phasing problem could not be resolved by MR. The statistics of the processed data are summarized in Table 3.

Table 3: Summary of diffraction data for MDA5 C915S/C951S with 12 bp dsRNA

	MDA5 C915S/C951S and 12 bp dsRNA complex
Wavelength (Å)	1.28255
Resolution range (Å)	50.00-3.20 (3.26-3.20)
Space group	$P3_1$ or $P3_2$
Unit cell parameters (Å)	$a = b = 149.053$, $c = 73.428$
Total No. of reflections	853748
No. of unique reflections	76923
Completeness (%)	97.1 (88.9)
Mean I/σ (I)	27 (1.73)
R_{sym}[†] (%)	9.4 (77.5)

[†]$R_{sym} = \sum_h \sum_i |I_{i,hkl} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_i |I_{i,hkl}|$, where $I_{hkl,i}$ is the intensity measured for a given reflection with Miller indices h, k, and l, and $\langle I_{hkl} \rangle$ is the mean intensity of that reflection.

CHAPTER III

STRUCTURAL CHARACTERISTICS OF THE *E.coli* PROLINE PEPTIDASE PepQ

Introduction

Of all the 20 common amino acids, proline is the only residue with a structure that forms a cyclical bond with its amide backbone. This limits catalytic activity at the proline, as the backbone near the residue is less susceptible to proteolytic degradation than other residues. [27] As a result, many organisms have evolved multiple ways to specifically target and digest proline-containing peptides. [28] One pair of proline-specific catalytic enzymes in particular in *E. coli* is PepP (proline aminopeptidase) and PepQ (prolidase). These proteins have a high sequence similarity to the *Alteromonas sp.* organophosphorus acid anhydrolase (OpaA), which has both proline dipeptidase activity and is capable of hydrolyzing toxic organophosphorus compounds found in nerve agents. PepQ in particular has 51% sequence identity to OpaA and was also found to have strong hydrolysis activity for organophosphorus compounds. [29]

In solution, *E. coli* PepP has been characterized as forming a tetramer and hydrolyzes the peptide backbone just before the proline in the N-terminal motif NH_2-Xaa-Pro-Xaa-. In contrast, *E. coli* PepQ forms a dimer in solution and instead recognizes dipeptides with the motif NH_2-Xaa-Pro-CO_2H. [29] Of these two enzymes, the PepP structure (PDB ID: 1AZ9, 1JAW, 1M35) has already been determined. [30] Determination of the structure of PepQ would permit a comparison to PepP and similar homologues, and may elucidate why the oligomeric structure is different.

Specific aim: X-ray crystallography of PepQ

Purified full length PepQ protein was screened in various crystallization conditions and then diffracted by the Raxis IV home source. Data were collected and the unrefined crystal structure was resolved. Analysis of the structure and comparison to the structures of two PepQ homologues (PepQ from *Alteromonas macleodii* (3RVA) and OpaA from *Alteromonas sp.* (3L24, 3L7G)) is in progress.

Experimental Procedures

Sample preparation and crystallization of PepQ

A sample of the PepQ protein was provided by the Hays Rye laboratory in 5 mM DTT, 5 mM $MgCl_2$, 50 mM Tris, pH 7.5. Crystallization screening conditions were obtained from Hampton Research (USA): Index, Crystal Screen, Crystal Screen 2, and PEGRx-1. The protein was screened by the hanging drop method at 6 and 12 mg/mL. Drops composed of 2 μL sample and 2 μL condition were allowed to crystallize at 4°C. Irregular plate-like crystals formed after 10 days in 20% PEG MME 5000, 0.1 M Bis-Tris pH 6.5. Optimization of the crystals was performed and an optimized crystal was grown to 2.8 mm x 2.8 mm x 0.3 mm over 14 days. This crystal was transferred stepwise into a cryobuffer containing 30% PEG 400, 20% PEG MME 5000, 0.1 M Bis-Tris, pH 6.5. The crystal was mounted on a cryoloop and then flash-frozen in $N_2(l)$.

Data collection and structural analysis

Diffraction data for the optimized PepQ crystal were collected using a Raxis IV++ detector mounted on a Rigaku Micromax-007HF generator at 120 K. The crystal

was rotated through a total of 180 frames with 1° oscillations and 5 minutes per frame. The collected data were then analyzed and processed by the HKL2000 package. [22] The phasing problem was resolved by MR using the *Alteromonas macleodii* OpaA protein structure (3RVA) as a base structure. MR and refinement were performed using the Phenix package. [23]

Results

PepQ crystallization and data collection

The PepQ protein was provided by the Hays Rye laboratory at the Texas A&M University as a 12 mg/mL -80°C glycerol stock sample. To improve the crystallization potential of the protein sample, glycerol was removed prior to crystallization. An aliquot of the protein was diluted to 6 mg/mL and then both samples were crystallized as described. Protein crystals did not visibly form until at least 8 days of incubation at 4°C,

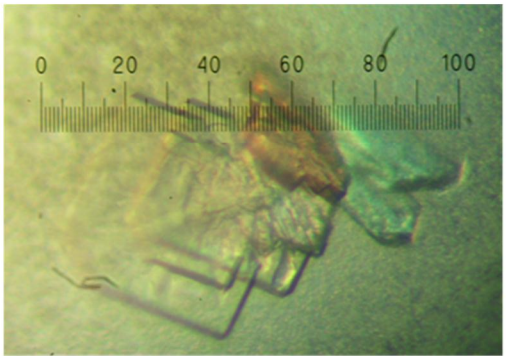

Fig. 16: Optimized PepQ crystal. Crystal shown formed on day 10 and was harvested on day 15. Scale is 63 μM per unit.

but the crystal growth rate quickly plateaued, with crystals reaching their maximum size two weeks post-screening.

Table 4: Statistics of diffraction data and refinement of *E. coli* PepQ

Data Collection		Refinement	
Wavelength (Å)	1.54178	R_{work}/R_{free}‡ (%)	20.1/25.7
Resolution range (Å)	30.20-2.26 (2.34-2.26)	Bond lengths (Å)	0.0160
Space group	$P2_12_12_1$	Bond angles (°)	1.58
Unit cell parameters (Å)	a = 71.111, b = 95.682, c = 127.329	Ramachandran plot (%)	
Total No. of reflections	853748	Favored	96.81
No. of unique reflections	76923	Allowed	2.62
Completeness (%)	98.9 (97.5)	Disallowed	0.57
Mean I/σ (I)	14.7 (3.2)		
R_{sym}† (%)	16.4 (64.0)		

†$R_{sym} = \sum_h\sum_i |I_{i,hkl} - \langle I_{hkl}\rangle|/\sum_{hkl}\sum_i |I_{i,hkl}|$, where $I_{hkl,i}$ is the intensity measured for a given reflection with Miller indices h, k, and l, and $\langle I_{hkl}\rangle$ is the mean intensity of that reflection.
‡$R_{work} = \sum||F_o|-|F_c||/\sum|F_o|$, where F_o and F_c are the observed and calculated structure-factor amplitudes, respectively. R_{free} was calculated as R_{work} using a randomly selected subset (10%) of unique reflections not used for structure refinement.

The *E. coli* PepQ protein was experimentally determined by the Hays Rye laboratory to be inhibited by HEPES, and so the addition of HEPES was used as a part of the crystallization optimization process. The addition of HEPES to the crystallization conditions impeded the nucleation process, and no crystals formed. The optimal

condition was found to be 19% PEG MME 5000, 0.1 M Bis-Tris, pH 6.5. The optimized crystal (Fig. 16) was grown as described and prepared for X-ray diffraction. Data was collected as described.

Data processing and structural analysis of PepQ

The *Alteromonas macleodii* OpaA protein has a sequence identity to the *E. coli* PepQ protein of 51%. The crystal structure of OpaA (3RVA) was used to develop a model PDB for PepQ using SWISS-Model. [15, 16, 31] This model structure was then used to resolve the phasing problem by MR using the Phenix package. [23] The attempt was successful, and the unrefined structure is presented here. Statistics for the crystal structure are listed in Table 4.

The *E. coli* PepQ structure (Fig. 17) shows high similarity to the structures of *Altermonas macleodii* OpaA (3RVA) and *Altermonas sp.* OpaA (3L24), with a an aligned monomer-monomer RMSD of 0.63 Å and 0.65 Å, respectively. The unrefined electron density map of *E. coli* PepQ shows density within the active sites (Fig. 18). The active sites of both the *Altermonas macleodii* OpaA and *Altermonas sp.* OpaA homologs contain a pair of divalent cations (Mn^{2+}). It is likely that this electron density represents the expected pairs of divalent cations.

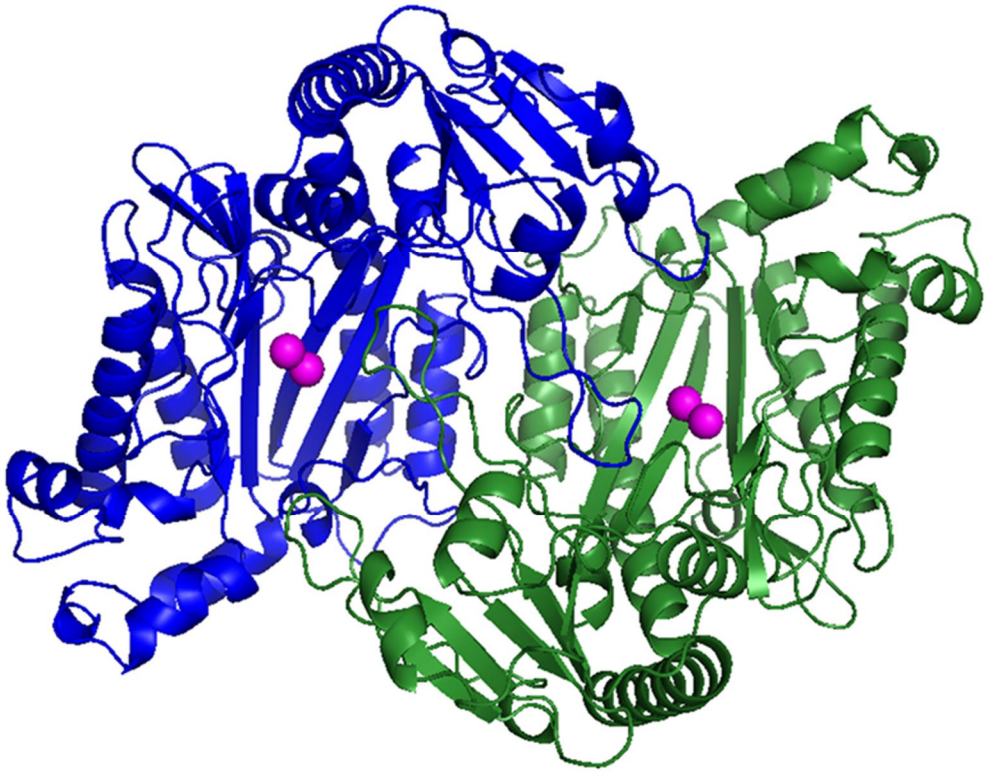

Fig. 17: PepQ crystal structure overview. The PepQ complete dimer with predicted divalent cation pairs placed as magenta spheres.

Of interest in this structure is that the dimer interface is explicitly characterized. The structure shows several key polar interactions that stabilize the symmetrical dimer interface: Asp45 with His346 and Arg370, Gln344 with Asp44 and the carbonyl of Leu43, and the carbonyl of Trp89 with the amide backbone of His228 and Leu227. The structure also shows three hydrophobic interfaces that lock the dimer in position (Fig. 19): a symmetrical core and two pockets positioned adjacent to the active sites of the dimer. The Trp89 reside is notable in that its carbonyl interacts with the amide backbone of His228, which is positioned along the outer edge of the active site of PepQ. (Fig. 20)

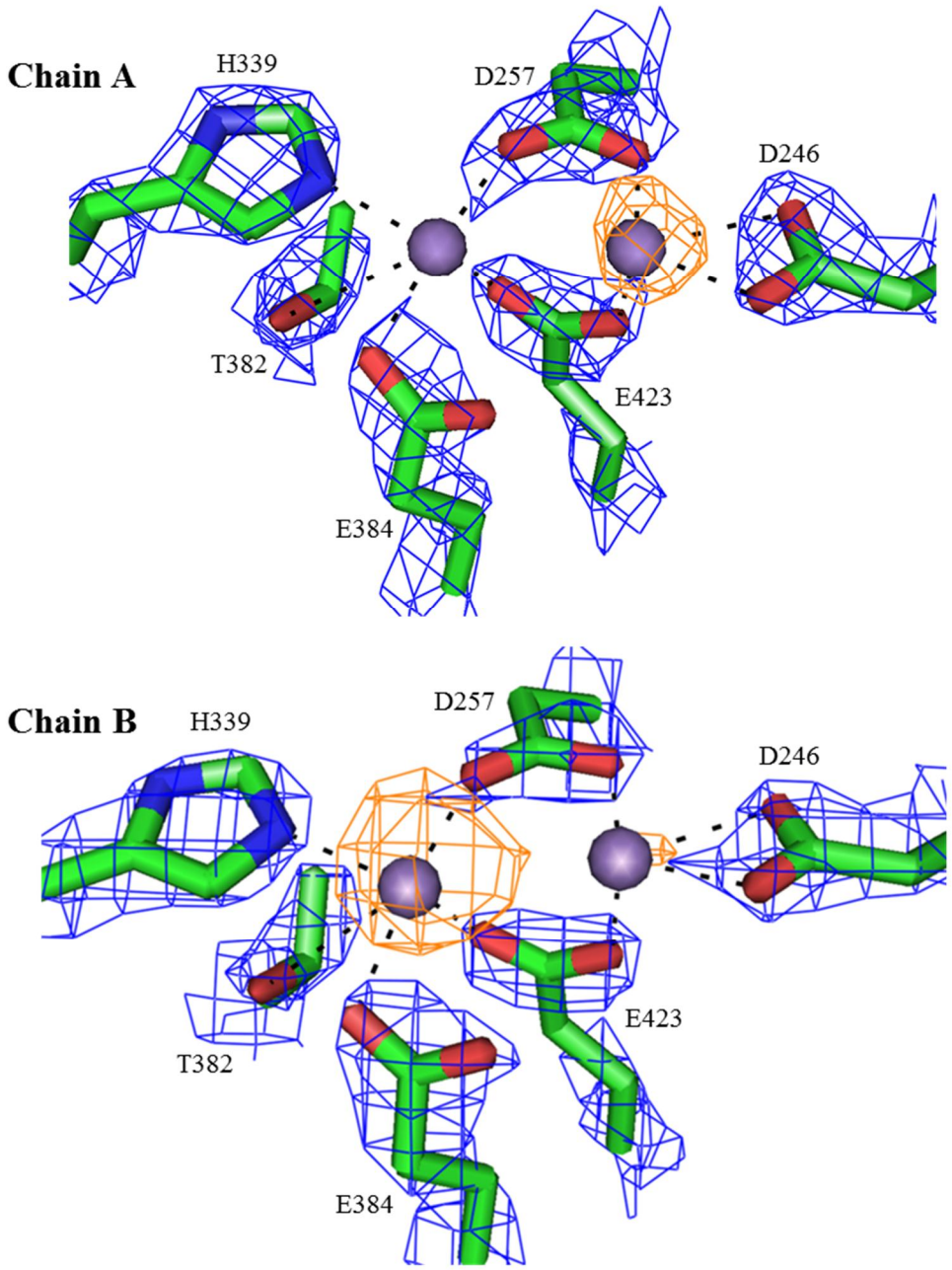

Fig. 18: PepQ active site. The active site of PepQ chain A and chain B showing interactions (black dashed lines) with divalent cation pair (purple spheres); the 2Fo-Fc (blue mesh) and Fo-Fc (orange mesh) are shown.

It may be that once the Trp89 is involved with the hydrophobic dimer interface, the carbonyl of Trp89 pulls the His228 residue away from the active site, helping to form the outer edge of the active site pocket.

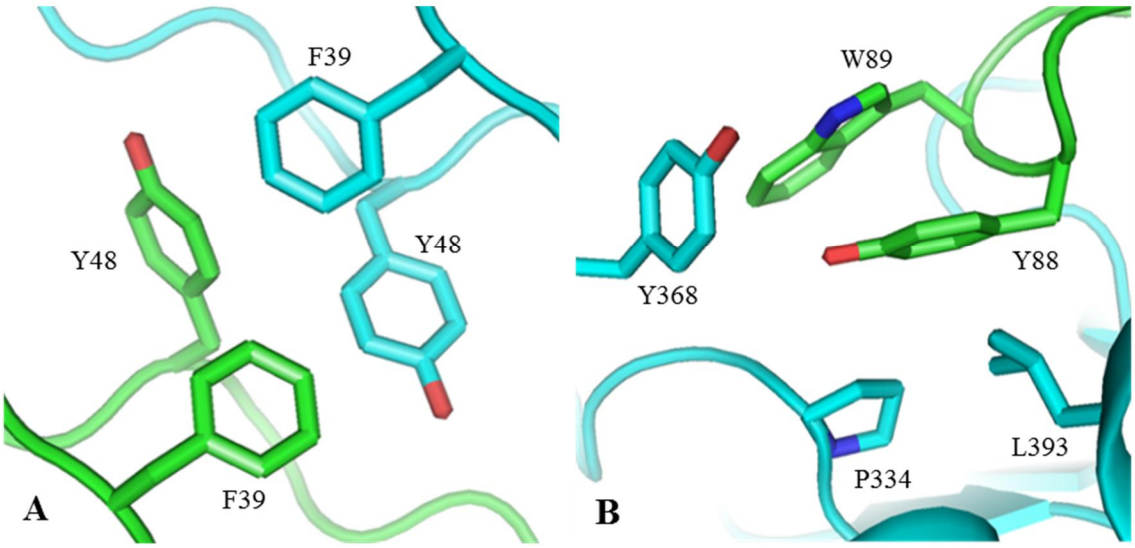

Fig. 19: Hydrophobic regions of the PepQ dimer interface. A: Central core of the dimer interface showing the positions of Phe39 and Tyr48 from both chains. B: hydrophobic region positioned near the active site of one chain; a similar interaction is seen near the active site of the other chain.

Further refinement of this protein structure will need to be performed to properly position the divalent cations. Also, the validity of the interactions shown here should be tested experimentally, especially in the case of Trp89. If the conformation of Trp89 and associated interaction with His226 is indeed essential for the formation of the active site, this would provide valuable insight into how the inactive PepQ monomer becomes an active dimer.

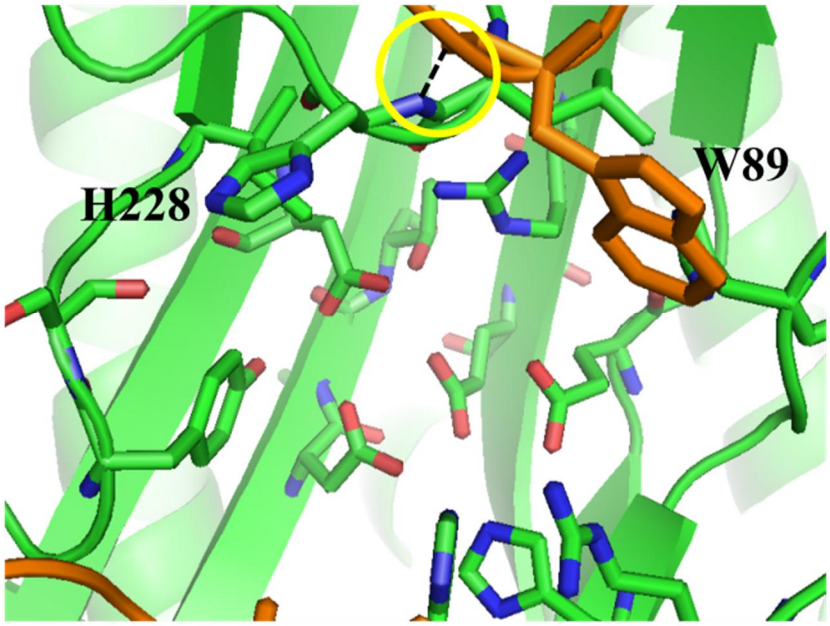

Fig. 20: Trp89 involved in the activation of the PepQ dimer. The carbonyl of the chain A Trp89 helps position the chain B H228 by a carbonyl-amide interaction.

CHAPTER IV

SUMMARY

Within this publication, both *H. sapiens* MDA5 and *E. coli* PepQ were studied. Experiments involving various techniques were performed on MDA5 constructs expressed from *E. coli,* with mixed levels of success. It was determined that the MDA5 Cys951 residue, localized on a flexible loop near the MDA5 RNA binding surface, has an influence upon the binding of short dsRNA. In addition, MDA5 binding to the 5' end of dsRNA is impaired by the presence of a 5' triphosphate group. The effect of 5' overhangs and 3' overhangs on the binding of MDA5 to dsRNA is unclear. The HD of MDA5 does not directly bind to RNA, but its influence on binding within the full length protein is likewise not certain. These results support the currently established data on this particular protein.

Studies on PepQ were restricted to structural analysis by X-ray crystallography. The purified protein was crystallized, diffraction data was collected, and the structure was successfully solved as a dimer. The active site for each subunit was found to relate very closely with the structures of *Altermonas macleodii* OpaA (3RVA) and *Altermonas sp.* OpaA (3L24). Three regions along the dimer interface show symmetrical hydrophobic interactions that lock the dimer, with a potential point of influence on the active site of one subunit by the Trp89 of the other subunit. The information from this structure will be used to further explore the nature and functions of the proline peptidases of bacteria.

REFERENCES

1. Kumar, H., T. Kawai, and S. Akira, *Pathogen recognition by the innate immune system.* Int Rev Immunol, 2011. **30**(1): p. 16-34.

2. Lightfield, K.L., et al., *Critical function for Naip5 in inflammasome activation by a conserved carboxy-terminal domain of flagellin.* Nat Immunol, 2008. **9**(10): p. 1171-8.

3. Shu, C., et al., *Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system.* Nat Struct Mol Biol, 2012. **19**(7): p. 722-4.

4. Franchi, L. and G. Nunez, *AIM2 joins the gang of microbial sensors.* Cell Host Microbe, 2010. **7**(5): p. 340-1.

5. Pichlmair, A. and C. Reis e Sousa, *Innate recognition of viruses.* Immunity, 2007. **27**(3): p. 370-83.

6. Akira, S., S. Uematsu, and O. Takeuchi, *Pathogen recognition and innate immunity.* Cell, 2006. **124**(4): p. 783-801.

7. Lee, M.S. and Y.J. Kim, *Signaling pathways downstream of pattern-recognition receptors and their cross talk.* Annu Rev Biochem, 2007. **76**: p. 447-80.

8. Besch, R., et al., *Proapoptotic signaling induced by RIG-I and MDA-5 results in type I interferon-independent apoptosis in human melanoma cells.* J Clin Invest, 2009. **119**(8): p. 2399-411.

9. Kowalinski, E., et al., *Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA.* Cell, 2011. **147**(2): p. 423-35.

10. Wu, B., et al., *Structural Basis for dsRNA Recognition, Filament Formation, and Antiviral Signal Activation by MDA5.* Cell, 2013. **152**(1-2): p. 276-89.

11. Li, X., et al., *Structural basis of double-stranded RNA recognition by the RIG-I like receptor MDA5.* Arch Biochem Biophys, 2009. **488**(1): p. 23-33.

12. Berke, I.C. and Y. Modis, *MDA5 cooperatively forms dimers and ATP-sensitive filaments upon binding double-stranded RNA.* EMBO J, 2012. **31**(7): p. 1714-26.

13. Luo, D., et al., *Visualizing the determinants of viral RNA recognition by innate immune sensor RIG-I.* Structure, 2012. **20**(11): p. 1983-8.

14. Lu, C., et al., *The structural basis of 5' triphosphate double-stranded RNA recognition by RIG-I C-terminal domain.* Structure, 2010. **18**(8): p. 1032-43.

15. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.

16. Bordoli, L., et al., *Protein structure homology modeling using SWISS-MODEL workspace.* Nat Protoc, 2009. **4**(1): p. 1-13.

17. Luo, D., et al., *Structural insights into RNA recognition by RIG-I.* Cell, 2011. **147**(2): p. 409-22.

18. Guemuer, Y. *Combinaison de classifieurs statistiques, Application a la prediction de structure secondaire des proteines*. [PhD Thesis]. Université de Paris: Paris, France (1997).

19. Cole, C., J.D. Barber, and G.J. Barton, *The Jpred 3 secondary structure prediction server.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W197-201.

20. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices.* J Mol Biol, 1999. **292**(2): p. 195-202.

21. Buchan, D.W., et al., *Protein annotation and modelling servers at University College London.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W563-8.

22. Otwinowski, Z. and W. Minor, *Processing of X-ray Diffraction Data Collected in Oscillation Mode*, in *Methods in Enzymology*, C.W. Carter, Jr. and R.M. Sweet, Editors. 1997, Academic Press: New York. p. 307-326.

23. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

24. Li, P., Li, X., Lu, C., *Crystal structure of the C-terminal domain of human MDA5.* PDB ID: 3GA3, Feb 2009. URL: http://dx.doi.org/10.2210/pdb3ga3/pdb.

25. Peisley, A., et al., *Cooperative assembly and dynamic disassembly of MDA5 filaments for viral dsRNA recognition.* Proc Natl Acad Sci U S A, 2011. **108**(52): p. 21010-5.

26. Kato, H., et al., *Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5.* J Exp Med, 2008. **205**(7): p. 1601-10.

27. Bromme, D., et al., *Enzyme-substrate interactions in the hydrolysis of peptide substrates by thermitase, subtilisin BPN', and proteinase K.* Arch Biochem Biophys, 1986. **244**(2): p. 439-46.

28. Cunningham, D.F. and B. O'Connor, *Proline specific peptidases.* Biochim Biophys Acta, 1997. **1343**(2): p. 160-86.

29. Park, M.S., et al., *Catalytic properties of the PepQ prolidase from Escherichia coli.* Arch Biochem Biophys, 2004. **429**(2): p. 224-30.

30. Wilce, M.C., et al., *Structure and mechanism of a proline-specific aminopeptidase from Escherichia coli.* Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3472-7.

31. Stepankova, A., et al., *Organophosphorus acid anhydrolase from Alteromonas macleodii: structural study and functional relationship to prolidases.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2013. **69**(Pt 4): p. 346-54.